

MIXED REALITY CONTENT ALIGNMENT IN MONUMENTAL ENVIRONMENTS

S. Teruggi^{1*}, F. Fassi¹

¹3D Survey Group, ABC Dep, Politecnico di Milano, Via Ponzio 31, 20133, Milano, Italy.
(simone.teruggi, francesco.fassi)@polimi.it

Commission II, WG II/8

KEYWORDS: HoloLens, Mixed Reality, World Locking Tools, Vuforia, Alignment, Point Cloud.

ABSTRACT:

Mixed reality provides on-the-spot and real-time data access capabilities by making virtual models and information more intuitive and accessible. Furthermore, allowing the operator to perceive 3D as holograms would allow for a more natural and straightforward manipulation of the perceived 3D content by permitting the augmentation of real objects with various levels of data. This can be accomplished by appropriately registering and superimposing the presented 3D models with the surrounding environment. This work aims to provide a quantitative evaluation of HoloLens 2 capabilities in registering virtual content inside monumental spaces. Two different methodologies are evaluated: Vuforia image targets and Microsoft World Locking Tools (WLTs). Tests have been performed inside Milan Cathedral's monumental spaces. Here, ambience dimensions, single architectural element repetition and non-uniform lighting conditions push out-of-the-box methods to their limits. Results show that WLTs with their space pins API can correctly reference virtual content keeping deviations in the order of 15 cm coping with the scale error produced from sensors' drifts.

1. INTRODUCTION

Since always professionals in building construction or conservation used to only work on "a two-dimensional world" and only used 3D models to convey particular concepts. Only recently have they begun to feel the need to use the three-dimensional model more proficiently as a tool for collecting and sharing all data related to the building's life cycle, from design to maintenance (Fassi et al., 2011). As a result, the 3D virtual representation is now known as a "digital twin", emphasizing the fact that it is used to connect data, information that may define behaviour and even non-visible aspects of the thing. The same concept, normally used for new constructions, can be used also for Cultural Heritage (CH) (Fassi et al., 2015). Producing a thorough digital twin when working with unique and complicated buildings is difficult both geometrically and in terms of information. Even more difficult is determining how to apply the accumulated information successfully in real-world situations throughout time.

Considering only the geometrical investigations, the state of the art in CH surveying includes a variety of non-invasive techniques that collect 3D point clouds that accurately represent the 3D geometry of even the most difficult item. However, an enormous amount of post-processing work by expert operators is still required, today, to use these data in professional practice in CH, producing the necessary 2D or 3D elaborations that serve as a reference for yard activities. There is a growing desire for innovative ways to use the information, not just geometric, to improve building comprehension and decision-making, speed up communication between practitioners, and allow data administration. Complex Heritage buildings are distinguished by constantly active yard activities dedicated to their maintenance and preservation. These activities generate an ever-increasing amount of data that is challenging to manage. Architects, restorers, historians, archivists, installers, and decision-makers could benefit from a full, detailed, and inventive representation that can classify, collect, preserve, and reference information. BIM models, theoretically, now enable the possibility to

reference and archive all data connected to preservation activities and periodic inspections (Bruno and Roncella, 2018) but there is a need to find an easier and more direct way to use geometric and non-geometric information directly during the day by day activity and possibly on the field. This would allow for coordinated storage and management of historical data, easy analysis and querying, flexibility and information sharing (Bruno and Roncella, 2019). The topic of using 1:1 content in the real world is a hot topic in fields such as industry and medicine (Avalle et al., 2019; Desselle et al., 2020). Augmented reality (AR) and mixed reality (MR) provide on-the-spot and real-time data access capabilities by making content more intuitive and accessible and, in the case of MR, directly connected to the real world. Furthermore, allowing the operator to perceive such data as holograms would allow for a more natural and straightforward manipulation of the perceived 3D content by permitting the augmentation of real objects with various levels of data.

The HoloLens 2 (Microsoft, 2020) is a cutting-edge MR headset. It allows augmenting virtual content to the environment in the form of holograms. The user can use their voice and hands to manipulate these 3D objects in the real world. This opens a whole new world of possibilities. Combining real and digital allows for increased knowledge, understanding, and collaboration within long-term projects and complex collaborations where a high degree of cooperation is required. This is a common scenario in the field of CH, which might benefit from the use of MR in both communication and professional practice. Locating information and recording actions referenced in space could be highly helpful in facilitating collaboration, during the inspection and maintenance process during the time, by developing a 1:1 scale information system immediately related to the object investigated and directly usable within the site's activities.

The case of Milan Cathedral is exceptionally iconic for maintenance and inspection interventions performed by the Veneranda Fabbrica del Duomo di Milano (VF). Having the possibility to find directly on the object past or hidden information during inspections or regular maintenance activity in

* Corresponding author

real-time and on-site would undoubtedly enhance the virtual understanding process and help the decision making. (Teruggi et al., 2021). Until now, all marble blocks have been directly inspected by a professional operator, who assesses their state of conservation, hazard, and level of risk. Manual signatures are required for intervention blocks, and technical data is recorded on paper tables. The acquired information is only afterwards transformed into its digital form and kept locally, posing the risk of data loss due to ageing or deterioration, or simply because they are forgotten, or their origin is completely neglected.

The main goal of the ongoing research is to solve all of these issues by developing a system that will allow the digital twin of the Cathedral to be brought back on site to reference, store, collect, and manage all information resulting from the systematic and planned interventions, allowing for continuous use and updating of the data on-site. Operatively it is necessary to have an exhaustive 3D digital twin of the Cathedral, adapt it for use on mobile devices and reference it correctly within the natural environment. The model derives from the intensive survey campaign carried out by the 3D Survey Group of Politecnico di Milano that produced a complete 3D point cloud of the whole monument (Achille et al., 2020). This point model might serve as the foundation for an information system that VF could use in the field, thanks to the HoloLens 2. This can be accomplished by appropriately registering and superimposing the presented 3D models with the surrounding environment. Such alignment must be pretty accurate to correctly reference the digital object units (marble blocks) in the real space, to consequently reference punctual data on the digital objects. This means the need to reference information on marble blocks with centimetre accuracy. This referencing process is possible because of the HoloLens 2's capability of simultaneously mapping its surroundings while localizing itself in space (Ungureanu et al., 2020). This is accomplished mainly through its depth sensor and its visible-light cameras and IMU. This capability has been extensively assessed (Hübner et al., 2020; Teruggi and Fassi, 2022). However, superimposing the virtual content on the physical space still poses a real challenge because a correct alignment between virtual and real objects is, de facto, impossible with no external constraints in case of a vast environment. This work aims at evaluating the reliability of different alignment methods in a CH monumental environment where spaces' dimensions, the repetition of the same architectonic elements and the lighting conditions make it challenging to define a unique and standard pipeline. The actual state of the art does not provide updated and complete quantitative evaluations of such alignment methods for large environments. Few existing examples of this type of evaluation are confined to small objects or single rooms environments. Hübner et al. (2018) evaluate the capability of the HoloLens device (first generation) to augment one room with a holographic model through a marker-based methodology. They proved the potential in this type of holographic augmentation. Holographic superimposition through external libraries has been evaluated using Vuforia image target libraries to display holograms in medical applications (Frantz et al., 2018). The research advances the state-of-the-art by assessing the ability to align virtual models to real-world objects and keep them in place using Vuforia image target and Microsoft world locking tools (WLTs) technologies. Finally, the two methodologies will be compared, highlighting related advantages and disadvantages.

2. HOLOGRAPHIC POSITIONING SYSTEMS

All 3D graphic applications use a cartesian coordinate system to define the position of virtual objects. An MR application must

consider both virtual and physical coordinate systems because virtual objects must be placed in the real environment.

The HoloLens device uses a virtual coordinate system projected to the physical world called a "spatial coordinate system" (Microsoft, 2022a). It expresses the coordinate values in meters. Two items put virtually two units apart in the application will appear rendered two meters apart in the actual world. The normal practice in many visualization programs today is to define one "absolute coordinate system" to which all coordinates are mapped. There is always a stable transformation that defines the relationship between points. If the objects are not moved, the relationship does not change. This method works well when the goal is rendering a completely virtual environment where all the geometry is known in advance.

In contrast, HoloLens 2 has a dynamic sensor-driven understanding of the world. The device gathers this knowledge in the form of a spatial map, a 3D mesh model of its surroundings (Teruggi and Fassi, 2022). This is necessary to place holograms in the natural environment, to enable real objects to occlude virtual content and to allow 3D models to interact with their surroundings. Placing all of the holograms in a single rigid coordinate system in this situation will cause them to drift over time, either around themselves or to the real world. Because the HoloLens 2 employs its depth sensor to measure the real world, this is the case. Each time the surroundings are scanned, these measurements are fine-tuned.

From a "positioning" point of view, HoloLens is "egocentric". Each time it is started, it creates an arbitrary coordinate system positioned to i) have the device at the axis origin, ii) have the z-axis pointing upwards, iii) have the x-axis directed frontally. At the same time, the application scene is shown with the coordinate system aligned to this arbitrary system. The position of the Cartesian axes of the scene (and thus the position of scene objects) concerning the real-world coordinate system is random: it depends on where the HoloLens is started. Two main tools can allow the HoloLens to re-orient and re-align virtual content on the natural world coordinate system consistently through different user sessions: i) Vuforia targets and ii) Microsoft WLTs together with their space pins API.

2.1 Vuforia Target

The Vuforia libraries provide different types of targets to orient the virtual models within the real-world coordinate system. The two main types are i) image targets and ii) area targets.

2.1.1 Image Targets are physical images positioned in the real world. The Vuforia Engine can detect them and keep track of their position using computer vision capabilities (Vuforia, 2020). These targets can be created from any source image in .JPG or .PNG format. The images are processed through the Vuforia website extracting meaningful features as sharp, spiked, and chiselled details in the image (e.g., a square contains four features, one for each corner, a circle contains zero features) (Figure 1a). The website allows exporting a software package that includes the image target and the extracted features. It can be imported into the developing environment and deployed with the application. It is possible to provide information to the device about the virtual model's position associated with the image target itself by using an image target.

The image target (virtual object) has to be manually positioned and oriented inside the virtual cartesian coordinate system during the system preparation. The virtual image target serves as a reference for all 3D objects in the scene (e.g. an object positioned to the right of the virtual image target will appear on the right of the physical image target).

The orientation of this target highly influences the superimposition results. Since it is manually adjusted, a small error in the virtual image target's rotation will obviously result in significant shifts as the distance from the target increases. When the camera of the device recognizes the physical copy of the image target in the real world, features extracted from the image target are compared and matched against the known target resource database deployed with the application. Therefore, all 3D models are roto-translated in the virtual coordinate system and displayed in the correct position (Figure 2a).

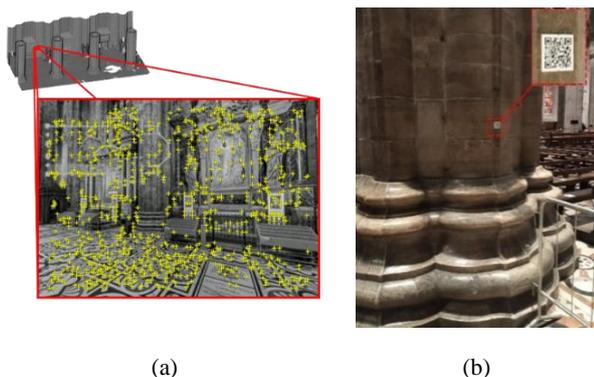


Figure 1. (a) Image target used in the south nave (A4 size) with features extracted using Vuforia target manager. (b) Space pin materialized as a QR code inside Milan Cathedral (5cm x 5cm).

2.1.2 Area Targets provide out of the box tracking and orientation of 3D models based on environment natural features. It is possible to generate an area target database that can be loaded into the development environment using a 3D scan as an exact model of a space. The HoloLens 2 camera continuously analyzes the environment, extracting information from it; these attributes are compared to those in the area target database, and augmented material is matched by orienting it in the real world. However, this target type can be generated only using Leica BLK360 and RTC360 terrestrial laser scanner or Matterport depth images. Those data types are not available for the Milan Cathedral and generally are not a standard for monumental heritage surveying. Therefore, they have been excluded from the following study.

2.2 Microsoft Content Alignment

Vuforia image target provides a powerful instrument to reference holograms in the real world, but this is based on one single static reference system. When the user moves through space, drifts due to sensor errors start to appear, and holograms move from the

desired positions. Microsoft allows to reference content in the real world through two different methods i) spatial anchors and ii) WLTs and their space pins API. While the firsts allow to tie the position of holograms to natural features (edges, salient points...) in the real world, the WLTs are an engine that offers different optimizations and improvements to HoloLens's way of referencing holograms using spatial anchors.

2.2.1 Spatial Anchors are virtual key points that tie the position of holograms to specific points of interest in the physical world. As the user moves from point A to B, errors in device position appear due to sensors drifts. Spatial anchors adjust their position by moving holograms inside the virtual coordinate system so that head coordinates are always right and the 3D models are displayed in the correct position in the physical space. As one spatial anchor changes the virtual coordinates of holograms that are tied to its position, it does so independently from where other anchors are in the physical space. The positions of holograms are refined point by point, in this way the model appears stable and well-aligned near the user, but not far away. A building site like that of Milan Cathedral cannot take full advantage of using spatial anchors. The Cathedral is characterized by huge dimensions and, using spatial anchors, drifts are not controllable in those areas (e.g., main naves) where the user can and wants to see far holograms as well in the correct position.

2.2.2 The World Locking Tools engine continuously spreads a supply of spatial anchors across the space as the user moves. In each frame, the WLTs engine inside the HoloLens calculates the coordinates of the camera position and those of the spatial anchors inside the virtual coordinate system.

When the engine detects that the spatial anchors are moving (to adjust the hologram's position at a particular point), it fixes the camera's virtual coordinates instead of changing the position of every other spatial anchor.

Instead of a spatial anchor dragging holograms around the virtual coordinate system to keep them anchored in physical space, the virtual coordinate system is locked to the real world. A hologram fixed in the virtual coordinate system will stay stationary in the real world. As important, it will remain fixed concerning other holograms.

When the user walks from point A to point B and then back again to point A, the WLTs engine recognize the initial point A position and adjusts the camera coordinates accordingly.

Still, one problem remains, the device, when moving from a point A to a distinct point B, 10 m apart, could register a 1 m error in the position of B, but it has no way of knowing if the registered 9 m distance is correct or not. The WLTs engine addresses the problem with their Space Pins API.

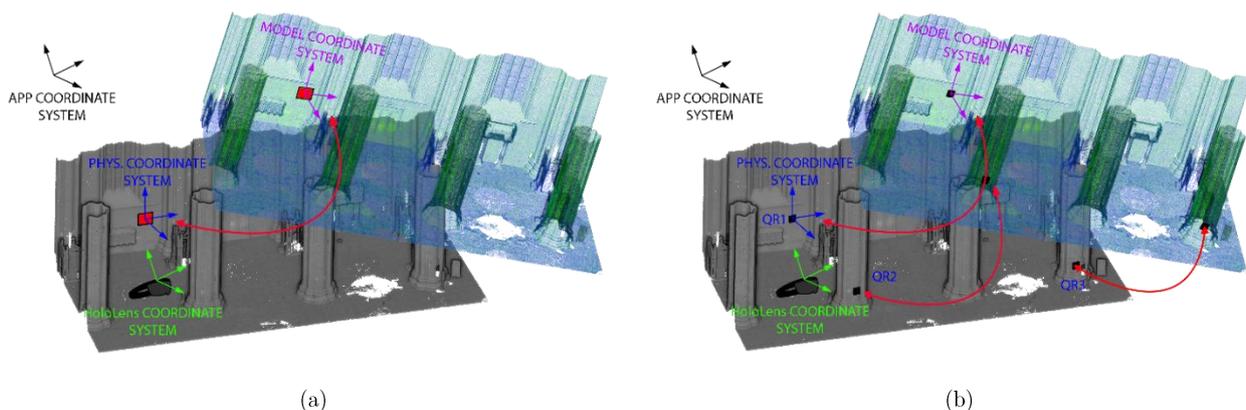


Figure 2. (a) Vuforia image target scheme. (b) Microsoft WLTs space pins API scheme.

The space pins API allow the application to supply known coordinates positions for different salient points. They allow for correctly aligning virtual content to the real world by picking salient points on the real object and moving the corresponding virtual points on the model in the correct position. This process could be automatized using physical QR codes that the device scan and recognize as their virtual counterparts (Figure 1b – 2b).

Using more than one space pin it is possible to supply to the application enough information regarding holograms distances in the real and holographic world to correct the scale errors. The process allows big holograms to appear aligned to the physical world from end to end. As the user moves between different space pins, a smooth interpolation minimizes the scale error at any given point in space.

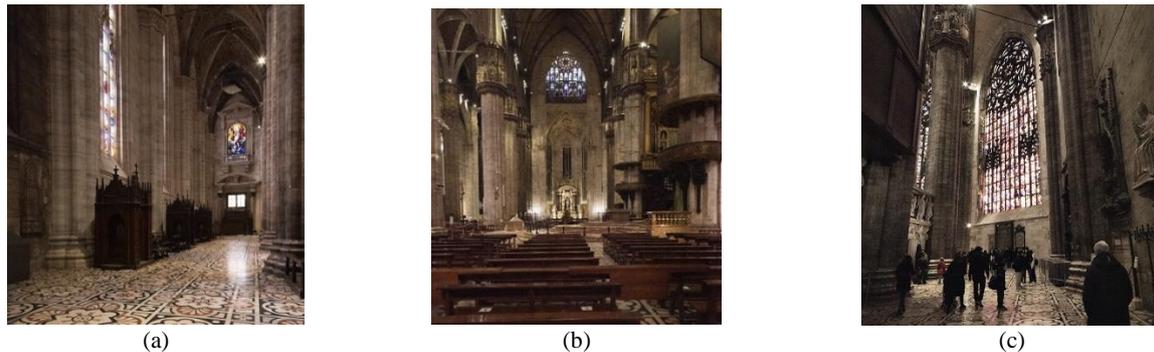


Figure 3. Milan Cathedral test areas. (a) South nave. (b) Transept (c) Apse.

3. TEST AREAS

Milan Cathedral constitutes the best example where to quantitatively assess the performance of Vuforia image targets and Microsoft WLTs space pins API. Its main noble areas (naves, transept and apse) are characterized by huge dimensions, architectural element repetition and non-uniform lighting conditions. All these factors directly affect the capability of the HoloLens 2 device to localize itself accurately and correctly inside the space. Furthermore, sensor drifts and errors that accumulate over time, as the user moves through space, are pushed to the limit and are well visible inside these kinds of spaces. Three different areas have been selected as test cases (Figure 3): i) the south nave; ii) the transept and iii) the apse. They have been selected considering the shape of the environment and its dimensions. The south nave area is very long, with a lot of single replicated architectural elements. It is defined by the repetition of a single bay module (9.5 x 9.5 x 24 m) eight times to form a massive space measuring 76 meters long, 9.5 meters wide, and 24 meters high. It is delimited on one side by a wall with large windows facing the exterior, and on the other side by a series of pillars. The transept constitutes the central transversal body of Milan Cathedral. It is divided into three naves, two smaller on the side (31 m height) and a central one (46 m height). The side naves are formed by the repetition of one bay module (9.5 x 9.5 x 31 m) while the central one repeats one double module (9.5 x 19 x 46 m) seven times. In the central part, where the main altar is located there is a big bay (19 x 19 x 65 m at its highest point) covered by the main dome under the tiburium on which the main spire rests. The total dimensions of the transept area are 40 x 85 m. The apse is constituted by the repetition of a trapezoidal bay module with dimensions of 14 m and 7 m for the long and short sides respectively and 9.5 m for the diagonal sides. It reaches 31 m at its highest point and has a semi-circular shape running around the choir of the Cathedral. Three big gothic windows illuminate the space from the east while the two sacristies are placed on the apse's sides.

4. METHODOLOGY

All test scenes have been set up inside Unity software (Unity, 2020). Vuforia image targets have been imported using the Vuforia Engine ver 10.5 (Vuforia, 2021). WLTs exploited the

capabilities included in ver. 1.5.8 (Microsoft, 2022b). Vuforia image targets have been tested positioning one A4 custom-designed image target for each test area. Measurements have been taken placing the image target at two different locations for each experiment. Firstly, the image target has been placed at one extreme of the area, then, the target has been positioned at the middle point. A total of 26 evenly distributed WLTs space pins have been materialized inside the Cathedral space as QR codes (5 x 5 cm in size). Each QR code has been positioned and measured with a total station. This allows knowing their precise location because referenced inside the Milan Cathedral topographic network. These coordinates have been used to correctly position their digital counterpart on the holographic point model to be displayed. The general idea is that as the HoloLens move through space it reads the different space pins (QR codes) and roto-translates the point model in the “real reference system”. In each test area, WLTs performances have been assessed by performing the acquisition in four different ways: i) scanning all QR codes present in the area; ii) scanning 3 QR codes in the initial part of the area; iii) scanning 3 QR codes evenly distributed; iv) scanning only 2 QR codes at the extremes of the considered portion. Drifts have been evaluated on 2 QR codes at the end of each area and its starting point after crossing the space backwards. Drifts at the initial point (before crossing the space) have been discarded since negligible. Errors are reported in terms of deviations in (x, y, z) coordinates expressed in m, in their absolute value. In all cases, drifts have been measured manually using a measuring tape to evaluate the deviations between the known location of the physical QR codes and their digital counterpart materialized as spheres holograms.

4.1 South Nave

4.1.1 Image Targets have been tested through two different target configurations. The first configuration has the target positioned in bay 1. The second one has it positioned in bay 5. Deviations have been measured by evaluating the distance between holographic spheres positioned at the space pin known coordinates and their real counterparts (QR codes positioned inside the south nave).

Measurements show that, as expected, image targets don't possess the ability to compensate for the scale error produced by sensor drift in the direction of movement.

In fact, deviations reach up to (0.98 m; 0.23 m; 0.11 m) on QR9 and (1 m; 0.22 m; 0.1 m) on QR10 with image target positioned in bay 1 (Figure 4a). With the image target positioned in bay 5

errors follows the same trend, yet, as the distance travelled from the target itself diminishes, so do the displayed errors (Figure 4b). However, deviations are never lower than 0.2 m.

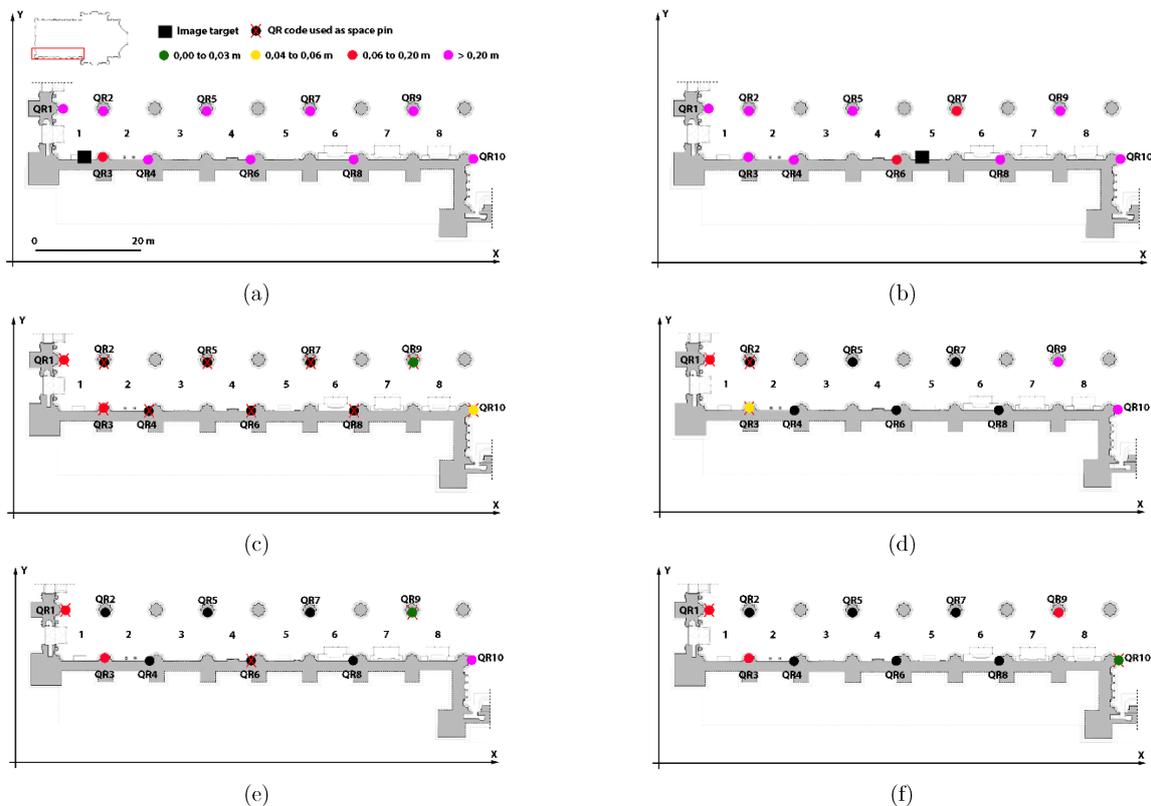


Figure 4. Image target deviations in the south nave (a) target positioned in bay 1; (b) target positioned in bay 5. Space pins API quantitative evaluation. The model has been roto-translated using (c) All 10 QR codes (d) QR 1 – 2 – 3. (e) QR 1 – 6 – 9. (f) QR 1 and 10. Deviations have been evaluated in bay 8 and, in bay 1 after crossing the space backwards.

4.1.2 WLTS' Space Pins API has been evaluated positioning a total of 10 QR codes inside the south nave. Four different QR code configurations have been tested: i) scanning all QR codes while crossing the space from bay 1 to bay 8 (Figure 4c); ii) scanning only 3 QR codes in bay 1 (QR1 – QR2 – QR3) (Figure 4d); iii) scanning 3 QR codes evenly distributed along the nave (QR1 – QR6 – QR9) (Figure 4e); iv) scanning 2 QR codes at the extremes (bay 1 and bay 8; QR1 – QR10) (Figure 4f). Deviations have been evaluated by measuring the distance between the real QR codes positioned inside the Cathedral and holographic spheres positioned at the measured coordinates inside Milan Cathedral topographic network.

Maximum deviations have been evaluated on QR9 and QR10 after crossing the space from bay 1 to bay 8 and on QR1 and QR3 (bay 1) traversing the space backwards. All QR codes have been scanned only during the first walk. WLTS' space pins API rely on the ability of the device in localizing itself inside the space. Inside the Cathedral their performance is affected by the dimension of the environment, the non-uniform lighting condition and the repetition of similar architectural elements. The device could lose track of its position in space, as it can frame only similar portions of the space, especially when the user makes sudden movements.

All QR codes configurations show the capability of space pins API in keeping content aligned to real objects. The biggest deviation can be appreciated when using only 3 QR codes in bay 1 to fix holograms' positions. In this case, the greatest deviation can be appreciated in the direction of the walk reaching (0.68 m; 0.3 m; 0.7 m) on QR9 and (0.9 m; 0.36 m; 0.8 m) on QR10. All other configurations keep deviations below 0.15 m.

4.2 Transept

4.2.1 Image Targets, in the transept area, have been tested i) placing the first image target in bay 1 (Figure 5a) and ii) placing the second one in bay 3 near the main altar (Figure 5b). The greatest deviation can be appreciated with the image target positioned on bay 1, at QR 26 it reaches (0.15 m; 0.04 m; 0.72 m). With the target placed at bay 3 near the main altar deviation on the same QR code is (0.56 m; 0.50 m; 0.45 m). In this case, as well it is possible to see that errors are due to two main reasons, the impossibility for the image target to cope with the deviations introduced by sensor drifts and, the errors deriving from the orientation of the virtual image target during development.

4.2.2 WLTS' Space Pins API accuracy has been evaluated positioning 8 QR codes acting as space pins in the transept area (from QR11 to QR14 and from QR23 to QR26). Four QR codes configurations have been evaluated: i) all QR codes scanned while crossing the space, from bay 1 to bay 7. Deviations have been evaluated on QR 25 and QR 26 and QR11 and QR12 after crossing the space backwards (bay 7 to bay 1) (Figure 5c); ii) only the four QR codes under the dome (bay 4) have been scanned (QR13 – 14 and QR23 – 24) and deviations have been evaluated on QR25 – 26 and QR11 – 12 (Figure 5d); iii) four QR codes at the extreme of the transept area have been scanned (QR 11 – 12 and QR 25 – 26) evaluating deviations on the four QR codes under the dome (QR 13 – 14 and QR 23 – 24) (Figure 5e); iv) Only two QR codes have been used as space pins at the opposite extremes of the transept (QR 12 and QR 25) (Figure 5f). The biggest deviations are appreciated while

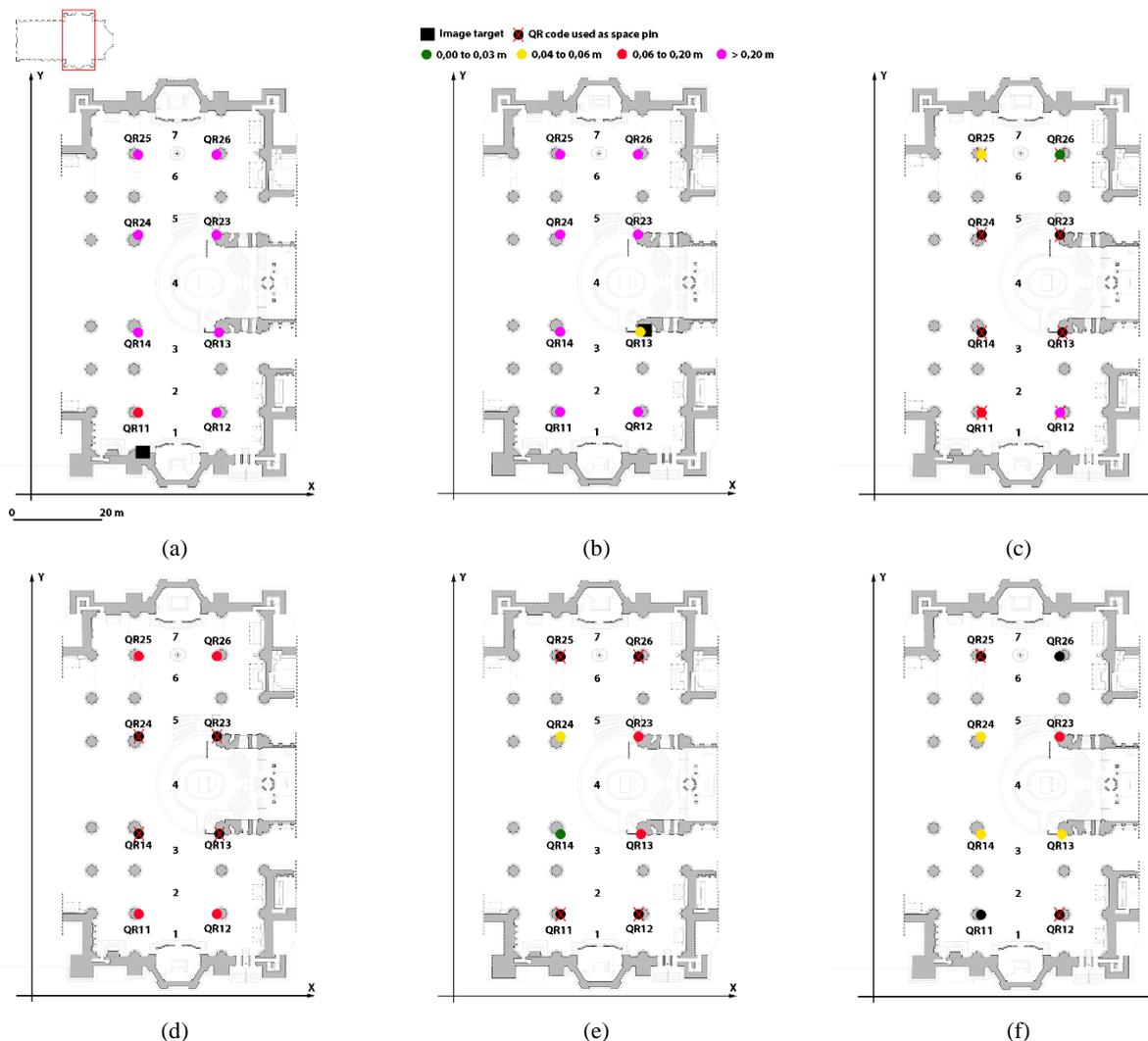


Figure 5. Image target deviations in the transept (a) target positioned in bay 1; (b) target positioned in bay 3 near the main altar. Space pins API quantitative evaluation. The model has been roto-translated using (c) All 8 QR codes. Deviations evaluated in bay 7 and bay 1, after crossing the space (d) QR13 – 14 – 23 – 24. Deviation evaluated on QR11 – 12 – 25 – 26 (e) QR 11 – 12 – 25 – 26. Deviations evaluated on QR 13 – 14 – 23 – 24 (f) QR 12 and QR25. Deviations evaluated on QR 13 – 14 – 23 – 24.

scanning only QR codes placed in bay 4 under the dome. Shifts at QR25 and QR26 reach (0.12 m; 0.01 m; 0.01 m) and (0.19 m; 0.02 m; 0.01 m) respectively. The trend is the same in the opposite direction on QR11 and 12 where distances between the QR codes and the displayed holographic sphere are (0.18 m; 0.02 m; 0.01 m) and (0.1 m; 0.05 m; 0.21 m). Different QR code configurations can keep deviations under 0.15 m with outliers in z-directions that reach up to 0.2 m. These are due to the difficulty that the device has in keeping track of its position in a so big and complex environment.

4.3 Apse

4.3.1 Image Targets have been tested using two different positions. The first test has been performed by scanning the image target positioned in bay 1 (Figure 6a), while the second one has been conducted with the target placed in bay 4 at the middle point of the apse area (Figure 6b). Deviations have been evaluated on all QR codes positions of known coordinates. As expected, these are higher than those obtained using WLTs space pins API. Distances from known points are (0.4 m; 0.2 m; 0.3 m) on QR22 with the target positioned in bay 1 and (0.48 m; 0.22 m; 0.28 m) on QR22 with the target positioned in bay 4.

4.3.2 WLTs' Space Pins API performance has been assessed using 8 QR codes (from QR 15 to QR 22). Four different QR codes configurations have been evaluated: i) all QR codes have been scanned and used as space pins while the area has been crossed from bay 1 to 7 (Figure 6c). ii) 3 QR codes positioned in bays 1 and 2 (QR15 – 16 – 17) have been used to register the position of the holograms (Figure 6d); iii) 3 evenly distributed QR codes (QR 15 – 18 – 22) have been scanned (Figure 6e) and iv) only 2 QR codes positioned in bay 1 and bay 7 have been used (QR 15 – QR 22) (Figure 6f).

Deviations, for all four different tests, have been evaluated on QR21 and QR22 after crossing the space the first time from bay 1 to 7 (scan of the QR codes is performed only during this walk) and on QR15 and QR16 after crossing the area backwards (no scan of the QR codes is performed during the walk). The highest deviations are measured, as expected, in bay 7 while using only 3 QR codes to register hologram positions in bay 1. They reach up to (0.25 m; 0.22 m; 0.06 m) on QR21 and (0.1 m; 0.35 m; 0.06 m) for QR22. Crossing the space backwards from bay 7 to 1, WLTs can contain sensor drifts accumulated during the walk, but an error of (0.14 m; 0.01 m; 0.05 m) and of (0.1 m; 0.04 m; 0.03 m) is visible on QR 15 and QR 16 respectively.

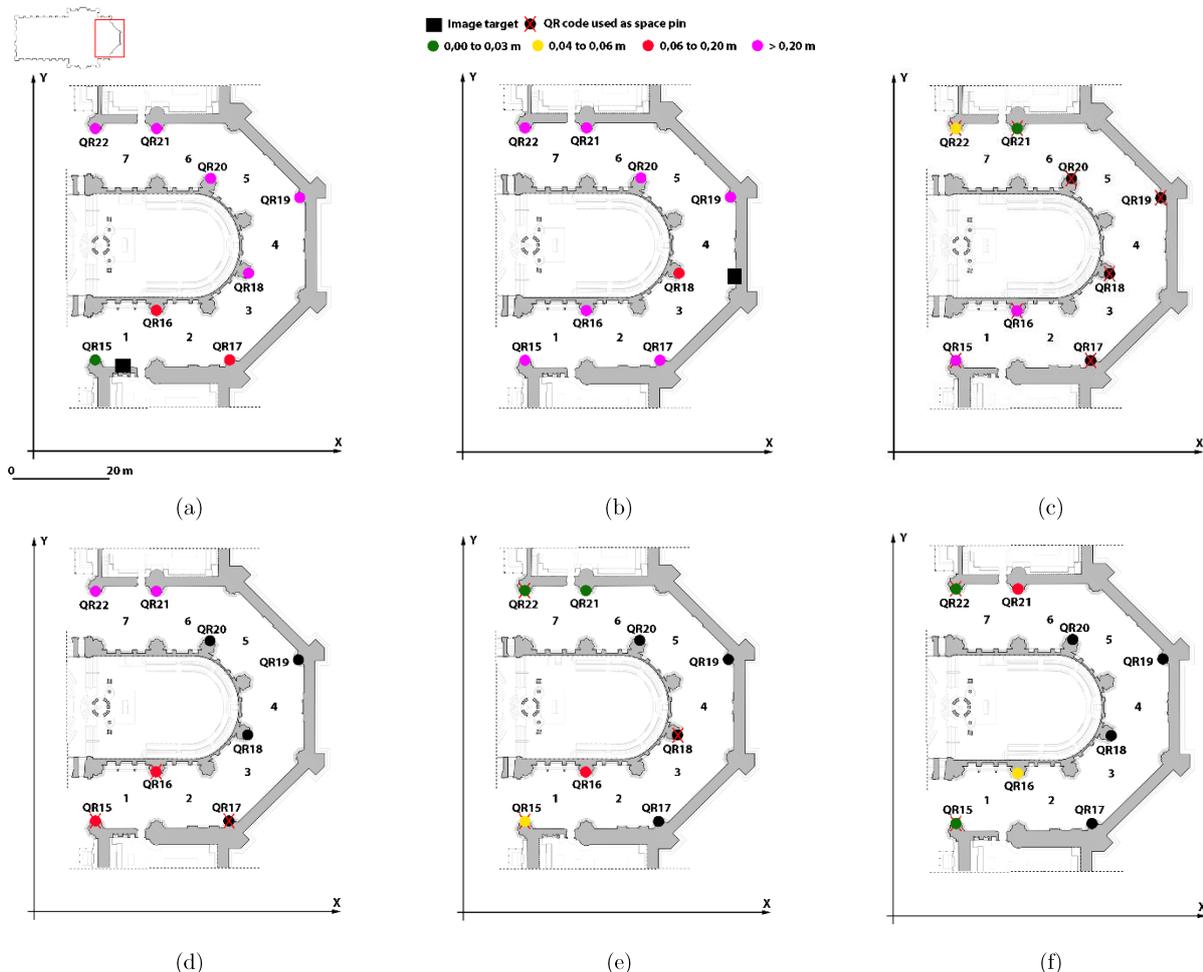


Figure 6. Image target deviations in the apse (a) target positioned in bay 1; (b) target positioned in bay 4. Space pins API quantitative evaluation. The model has been roto-translated using (c) All 8 QR. (d) QR 15 – 16 – 17. (e) QR 15 – 18 – 22. (f) QR 12 and 25. Deviations have been evaluated in bay 7 and, in bay 1 after crossing the space backwards.

Due to its semi-circular morphology, this part of Milan Cathedral proved the most challenging in which to keep holograms at stable positions. Deviations in bay 1, while using all QR codes positioned in the area, are in the order of 20/30 cm after crossing the space backwards from bay 7 to bay 1. When scanning only 3 QR codes in bay 1, errors on QR21 reach up to (0.25 m; 0.22 m; 0.06 m) and (0.10 m; 0.35 m; 0.06 m) on QR22.

5. DISCUSSION AND CONCLUSIONS

The work presented a quantitative evaluation of image targets and WLTs space pins API to align virtual content with the real world inside monumental heritage environments. Test performed proved the capabilities of both space pin API and Image targets. However, monumental spaces proved to be a real unresolved challenge for the process. Tests have been conducted inside three different areas of Milan Cathedral: i) the south nave, ii) the transept and iii) the apse. These typologies of spaces are characterized by huge dimensions, single architectural element repetition and non-uniform lighting conditions. Since the device relies on its visual-inertial slam and on its IMU sensor to keep track of its position inside the space, environmental conditions highly influence the accuracy with which such position is computed. The HoloLens 2 is exceptionally good to keep track of its position inside the space. However, as the user walks through space, errors produced by the device's sensors start to accumulate and holograms start to shift from their positions.

Image target and WLTs' space pin API provide two ways to register holographic content on the real-world coordinate system. The firsts use physical images at known locations about which holograms are positioned during application deployment. They provide one single rigid reference system that allows roto-translate 3D model virtual coordinates to match the position of the image target plane. As expected, they proved incapable of coping with the scale error generated by sensor drifts while moving the HoloLens through the space.

As shown in Figure 7, shifts increase with the distance from the target's position. This is for two main reasons: i) the stationary frame of reference does not allow tracking device errors that accumulate over time, and ii) rotations of the digital image target are manually set by the developer during deployment. It results that even small errors in orientation, negligible near the image target, result in a big shift while the distance is increasing. One example is the increasing error in z-direction inside the Transept area where on QR26 deviations reach up to (0.15 m; 0.04 m; 0.72 m). The Vuforia target's inability to cope with the scale inaccuracy caused by sensor drifts, on the other hand, is evident in the south nave, where errors could reach up to 1m on the farthest point from the target in the walking direction (Figure 4a). WLTs, on the contrary, thanks to their space pins API, allow them to orient the displayed virtual content using more points distributed in space as reference. This allows reaching higher accuracy in the position of the 3D model in the physical space. Furthermore, WLTs allow inserting measured coordinates of the QR codes registered from the Milan Cathedral surveys'

topographic network during app deployment. Knowing the exact coordinates of each space pin gives the possibility to compensate for the scale error produced by sensor drifts. Experiments showed that even with a small number of space pins distributed in space, deviations on QR codes are kept in the order of 15 cm.

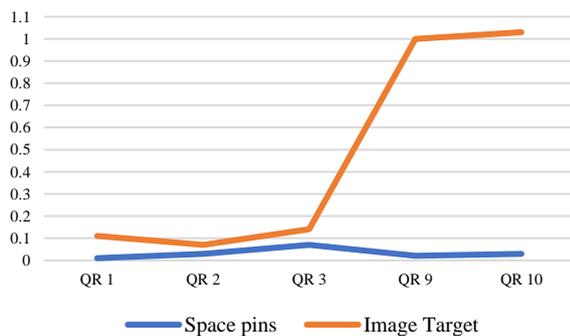


Figure 7. Deviation comparison in the south nave x-direction. The image target is positioned in bay 1, and all QR codes are used as space pins.

However, errors are distributed along with the distance that connects two different space pins. For this reason, overall alignment quality is better when using more QR codes distributed inside a huge ambient. From the obtained results, WLTs, thanks to their space pins API provide a more reliable, replicable and accurate solution to align holograms on real objects inside monumental spaces. However, Milan Cathedral's environment makes it impossible to exploit WLTs' full potential. At the moment, the device is unable to automatically re-orient the displayed material based on ambient circumstances once each session begins, and each application session necessitates scanning QR codes. This leaves open research questions where different future works are already in the program, from providing the device with the ability to better re-localize itself to better tracking its position inside these particular types of environments. In the end, WLTs offer the best out of the box solution to cope with environmental dimensions and conditions.

REFERENCES

Achille, C., Fassi, F., Mandelli, A., Perfetti, L., Rechichi, F., Teruggi, S., 2020. From a Traditional to a Digital Site: 2008–2019. The History of Milan Cathedral Surveys, *Research for Development*. https://doi.org/10.1007/978-3-030-33570-0_30

Avalle, G., Pace, F.D., Fornaro, C., Manuri, F., Sanna, A., 2019. An Augmented Reality System to Support Fault Visualization in Industrial Robotic Tasks. *IEEE ACCESS* 7, 132343–132359. <https://doi.org/10.1109/ACCESS.2019.2940887>

Bruno, N., Roncella, R., 2019. HBIM for Conservation: A New Proposal for Information Modeling. *Remote Sens.* 11. <https://doi.org/10.3390/rs11151751>

Bruno, N., Roncella, R., 2018. A Restoration Oriented HBIM System for Cultural Heritage Dissemination: The Case Study of Parma Cathedral. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII–2, 171–178. <https://doi.org/10.5194/isprs-archives-XLII-2-171-2018>

Desselle, M.R., Brown, R.A., James, A.R., Midwinter, M.J., Powell, S.K., Woodruff, M.A., 2020. Augmented and virtual reality in surgery. *Computing in Science and Engineering*. <https://doi.org/10.1109/MCSE.2020.2972822>

Fassi, F., Achille, C., Fregonese, L., 2011. Surveying and modelling the main spire of Milan Cathedral using multiple data sources. *Photogrammetric Record* 26, 462–487. <https://doi.org/10.1111/j.1477-9730.2011.00658.x>

Fassi, F., Achille, C., Mandelli, A., Rechichi, F., Parri, S., 2015. A New Idea of BIM System for Visualization, Web Sharing and Using Huge Complex 3d Models for Facility Management. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 40, 359–366. <https://doi.org/10.5194/isprsarchives-XL-5-W4-359-2015>

Frantz, T., Jansen, B., Duerinck, J., Vandemeulebroucke, J., 2018. Augmenting Microsoft's HoloLens with vuforia tracking for neuronavigation. *Healthc. Technol. Lett.* 5, 221–225. <https://doi.org/10.1049/htl.2018.5079>

Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., Wursthorn, S., 2020. Evaluation of hololens tracking and depth sensing for indoor mapping applications. *Sensors (Switzerland)* 20. <https://doi.org/10.3390/s20041021>

Hübner, P., Weinmann, M., Wursthorn, S., 2018. Marker-Based Localization of the Microsoft HoloLens in Building Models. *The Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII–1, 195–202. <https://doi.org/10.5194/isprs-archives-XLII-1-195-2018>

Microsoft, 2022a. Coordinate systems - Mixed Reality. <https://docs.microsoft.com/en-us/windows/mixed-reality/design/coordinate-systems> (accessed 3.26.22).

Microsoft, 2022b. Releases WorldLockingTools-Unity <https://github.com/microsoft/MixedReality-WorldLockingTools-Unity/releases> (accessed 3.28.22).

Microsoft, 2020. HoloLens 2 - Overview, funzionalità e specifiche. <https://www.microsoft.com/it-it/hololens/hardware> (accessed 3.26.22).

Teruggi, S., Fassi, F., 2022. HoloLens 2 Spatial Mapping Capabilities in Vast Monumental Heritage Environments. *The Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLVI-2/W1-2022, 489–496. <https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-489-2022>

Teruggi, S., Grilli, E., Fassi, F., Remondino, F., 2021. 3D Surveying, Semantic Enrichment and Virtual Access of Large Cultural Heritage. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* VIII-M-1-2021, 155–162. <https://doi.org/10.5194/isprs-annals-VIII-M-1-2021-155-2021>

Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., Stühmer, J., Cashman, T.J., Tekin, B., Schönberger, J.L., Olszta, P., Pollefeys, M., 2020. HoloLens 2 Research Mode as a Tool for Computer Vision Research *arXiv preprint arXiv:2008.11239*. <https://doi.org/10.48550/arXiv.2008.11239>

Unity, 2020. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. <https://unity.com/> (accessed 3.28.22).

Vuforia, 2021. Vuforia Developer Portal [WWW Document]. URL <https://developer.vuforia.com/downloads/SDK> (accessed 3.28.22).

Vuforia, 2020. Image Targets - Vuforia library. <https://library.vuforia.com/features/images/image-targets.html> (accessed 3.30.22).